

# Weiqi games as a tree: Zipf's law of openings and beyond

LI-GONG XU<sup>1</sup>, MING-XIA LI<sup>1,2,3</sup> and WEI-XING ZHOU<sup>3,4,5</sup> (a)

<sup>1</sup> School of Sports Science and Engineering, East China University of Science and Technology, Shanghai 200237, China

<sup>2</sup> Postdoctoral Research Station, East China University of Science and Technology, Shanghai 200237, China

<sup>3</sup> Research Center for Econophysics, East China University of Science and Technology, Shanghai 200237, China

<sup>4</sup> School of Business, East China University of Science and Technology, Shanghai 200237, China

<sup>5</sup> Department of Mathematics, East China University of Science and Technology, Shanghai 200237, China

PACS 01.80.+b – Physics of games and sports

PACS 89.20.-a – Interdisciplinary applications of physics

PACS 89.75.Da – Systems obeying scaling laws

PACS 89.75.Hc – Networks and genealogical trees

**Abstract** – Weiqi is one of the most complex board games played by two persons. The placement strategies adopted by Weiqi players are often used to analog the philosophy of human wars. Contrary to the western chess, Weiqi games are less studied by academics partially because Weiqi is popular only in East Asia, especially in China, Japan and Korea. Here, we propose to construct a directed tree using a database of extensive Weiqi games and perform a quantitative analysis of the Weiqi tree. We find that the popularity distribution of Weiqi openings with a same number of moves is distributed according to a power law and the tail exponent increases with the number of moves. Intriguingly, the superposition of the popularity distributions of Weiqi openings with the number of moves no more than a given number also has a power-law tail in which the tail exponent increases with the number of moves, and the superposed distribution approaches to the Zipf law. These findings are the same as for chess and support the conjecture that the popularity distribution of board game openings follows the Zipf law with a universal exponent. We also find that the distribution of out-degrees has a power-law form, the distribution of branching ratios has a very complicated pattern, and the distribution of uniqueness scores defined by the path lengths from the root vertex to the leaf vertices exhibits a unimodal shape. Our work provides a promising direction for the study of the decision making process of Weiqi playing from the angle of directed branching tree.

**Introduction.** – Weiqi, also called Go in Japan and Budak in Korea, originated over 3000 years ago in China and is probably the oldest board game in the world. The standard modern Weiqi game is played on a  $19 \times 19$  board. It is played by two players using black and white stones in turn. The rules of Weiqi playing are very simple. However, its strategies are very complex, much more than western chess. Specifically, Weiqi is PSPACE-hard [1] and EXPTIME-complete [2] and certain subproblems of the game are PSPACE-complete [3]. It has been shown that the state-space complexity of Weiqi is  $\sim 10^{171}$  and its game-tree complexity is  $\sim 10^{360}$  [4, 5]. For comparison, these two numbers are  $10^{47}$  and  $10^{123}$  for chess.

Although computer programs are able to defeat top-tier professional players for many mind board games including chess and Chinese chess (Xiangqi), the situation is completely different for Weiqi. It is even very hard for a computer to judge if a cluster of stones is alive or dead, and it is much harder to solve correctly a lot of life-and-death problems [6, 7]. Certainly, there are progresses on predicting life or death in the game of Weiqi through learning examples extracted from game records [8]. No doubt, the game of Weiqi has been a big challenge to researchers in the field of Artificial Intelligence (AI) [9–12]. Interestingly, in modern neuroscience, it has been shown that long-term trained Weiqi players “developed larger regions of white matter with increased fractional anisotropy values in the

(a) e-mail: wxzhou@ecust.edu.cn

frontal, cingulum, and striato-thalamic areas that are related to attentional control, working memory, executive regulation, and problem-solving” [13].

Recently, there are also efforts to study Weiqi games from the perspective of statistical physics. Liu, Dou and Lu defined an area, say a  $5 \times 5$  plaquette, centered by the current move and obtained a contextual pattern with successive moves within the plaquette. It is found that the occurrence distribution of contextual patterns follows Zipf’s law [14]. Harre et al. investigated the probabilities of different moves within the  $7 \times 7$  corner plaquette to quantify the lack of predictability of experts and how this changes with their levels of skill [15]. Georgeot and Giraud defined the 1107 plaquettes of size  $3 \times 3$  with empty centers as vertices and, if the distance between two successive moves is smaller than a preset value, the two associated vertices are connected with an arrow [16]. Alternatively, when the atari status of the four nearest neighbor points from the center is taken into consideration, one obtains 2051 legal nonequivalent plaquettes or vertices with empty centers [17, 18], and considering diamond-shape plaquettes by expanding the  $3 \times 3$  plaquettes and four additional points results in 193 995 nonequivalent plaquettes with empty centers [18]. Very intriguing results have been obtained for these networks [15–18].

In this Letter, we construct a directed Weiqi tree using a database of professional games, which is a subset of the whole strategy tree studied in the AI literature. The Weiqi tree constructed here adopts the same method as in Ref. [19] for chess. The method takes the move order into consideration so that the network is a tree and the in-degrees of all non-root vertices are 1. We will study the distributions of the popularity of Weiqi openings, the out-degrees, the branching ratios and the uniqueness scores of the Weiqi tree.

**Data description.** – The database of Weiqi games used in this Letter, called All-in-One Library of Weiqi Games (or “xún pǔ mǔ běn” in Chinese spell), is collected and maintained by the Gochess web site since more than 10 years ago, and can be downloaded from <http://www.gochess.cn/>. All the games were played by professionals. We use version 7.6 here, which has 69 031 games. After removing duplicate games, handicap games, transposon games, games on  $13 \times 13$  board, life or dead problems, non-human played games, and games with less than 30 moves, there are 57 090 games left for further analysis.

**Construction of the Weiqi tree.** – We denote each game by  $\mathcal{G}_i$ . For a Weiqi game  $\mathcal{G}_i$ , the  $k$ -th move can be determined by a point  $\mathcal{M}_k = (x_k, y_k)$  on the board, and the opening with  $m$  moves is a sequence of  $m$  points:  $\mathcal{G}_{i,m} = \{\mathcal{M}_k : k = 1, 2, \dots, m\}$ . The coordinates  $(x, y)$  of the  $19 \times 19$  board are shown in fig. 1.

Two main features of Weiqi games are rotation equivalence and flip equivalence, which means that two openings  $\mathcal{G}_{i,m}$  and  $\mathcal{G}_{j,m}$  overlap after certain rotation and/or flip.

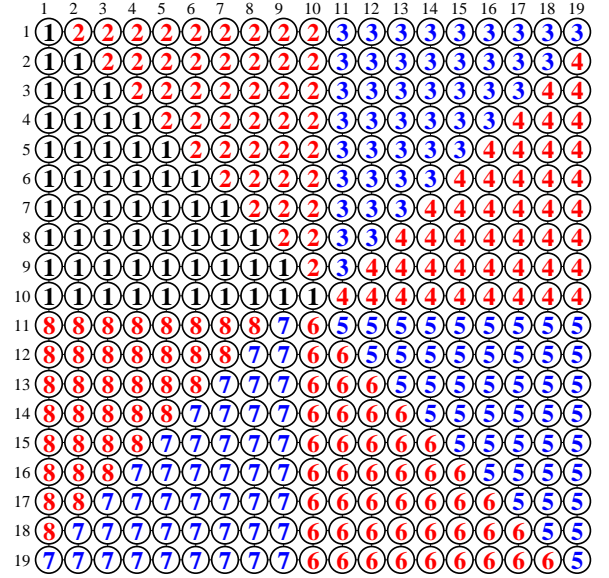


Fig. 1: Partitioning the Weiqi board into eight zones:  $Z_1 = \{x \leq y \leq 10\}$ ,  $Z_2 = \{y < x \leq 10\}$ ,  $Z_3 = \{x + y \leq 20, x > 10\}$ ,  $Z_4 = \{x + y > 20, y \leq 10\}$ ,  $Z_5 = \{10 < y \leq x\}$ ,  $Z_6 = \{x < y, x \geq 10\}$ ,  $Z_7 = \{20 \leq x + y, x < 10\}$ , and  $Z_8 = \{y + x < 20, x < 10, y > 10\}$ . The crosses in the first zone are determined. For other zones, it is optional to include the crosses at  $x = y$ ,  $x = 20 - y$ ,  $x = 10$  or  $y = 10$ . Two zones both with even or odd numbers have the possibility to overlap after rotation, while two zones with even and odd numbers may overlap after flip and rotation.

Equivalent rotations contain  $90^\circ$ ,  $180^\circ$  and  $270^\circ$  clockwise rotations. Mathematically, we have

$$\mathcal{G}_{i,m}(x, y) \equiv \begin{cases} \mathcal{G}_{i,m}(20 - y, x), & 90^\circ \\ \mathcal{G}_{i,m}(20 - x, 20 - y), & 180^\circ \\ \mathcal{G}_{i,m}(y, 20 - x), & 270^\circ \end{cases} \quad (1)$$

Equivalent flips can be conducted with respect to the  $x$  or  $y$  axis. Using  $y$ -flip, we have

$$\mathcal{G}_{i,m}(x, y) \equiv \mathcal{G}_{i,m}(20 - x, y). \quad (2)$$

Combining a rotation and a flip, there are three more Weiqi patterns that are equivalent:  $\mathcal{G}_{i,m}(20 - y, 20 - x)$ ,  $\mathcal{G}_{i,m}(x, 20 - y)$ , and  $\mathcal{G}_{i,m}(y, x)$ .

As shown in fig. 1, two openings  $\mathcal{G}_{i,m}$  and  $\mathcal{G}_{j,m}$  can be identical after proper rotation and slip. Hence, we preprocess the games to make them unique. We require that the first move locate in Zone 1 after suitable rotation and flip. If the first move is in Zone 1, i.e.,  $\mathcal{M}_1 \in Z_1$ , we do nothing. If  $\mathcal{M}_1 \in Z_2$ , we flip it along  $x = y$  (equivalently, rotate  $90^\circ$  clockwise and then flip along  $x = 10$ ), obtaining  $\mathcal{G}_i(y, x)$  from  $\mathcal{G}_i(x, y)$  for all games. If  $\mathcal{M}_1 \in Z_3$ , we obtain  $\mathcal{G}_i(y, 20 - x)$  through rotating  $\mathcal{G}_i(x, y)$  by  $90^\circ$  counterclockwise. If  $\mathcal{M}_1 \in Z_4$ , we obtain  $\mathcal{G}_i(20 - x, y)$  through flipping  $\mathcal{G}_i(x, y)$  along  $x = 10$ . If  $\mathcal{M}_1 \in Z_5$ , we obtain  $\mathcal{G}_i(20 - x, 20 - y)$  through rotating  $\mathcal{G}_i(x, y)$  by  $180^\circ$  counterclockwise. If  $\mathcal{M}_1 \in Z_6$ , we obtain  $\mathcal{G}_i(20 - y, 20 - x)$

through flipping  $\mathcal{G}_i(x, y)$  along  $x + y = 20$ . If  $\mathcal{M}_1 \in Z_7$ , we obtain  $\mathcal{G}_i(20 - y, x)$  through rotating  $\mathcal{G}_i(x, y)$  by  $270^\circ$  counterclockwise. If  $\mathcal{M}_1 \in Z_8$ , we obtain  $\mathcal{G}_i(x, 20 - y)$  through flipping  $\mathcal{G}_i(x, y)$  along  $y = 10$ . We stress that the transformation of coordinates is carried out for all moves, not only for the first move.

After the first step of transformation, the first move  $\mathcal{M}_1$  has been placed in Zone 1. If  $x_1 = y_1$  and  $x_2 > y_1$ , we flip the board along the diagonal  $y = x$  so that  $\mathcal{G}_i(x, y) \Rightarrow \mathcal{G}_i(y, x)$ . Moreover, if  $x_1 = y_1$ ,  $x_2 = y_2$ ,  $x_3 > y_3$ , we also flip the board along the diagonal  $y = x$  so that  $\mathcal{G}_i(x, y) \Rightarrow \mathcal{G}_i(y, x)$ . More generally, if the first  $k$  moves are on the diagonal and the  $k + 1$  move has  $x_{k+1} > y_{k+1}$ , we flip the board along  $x = y$ . Usually, Weiqi games with  $k \geq 3$  first moves on the diagonal are very rare.

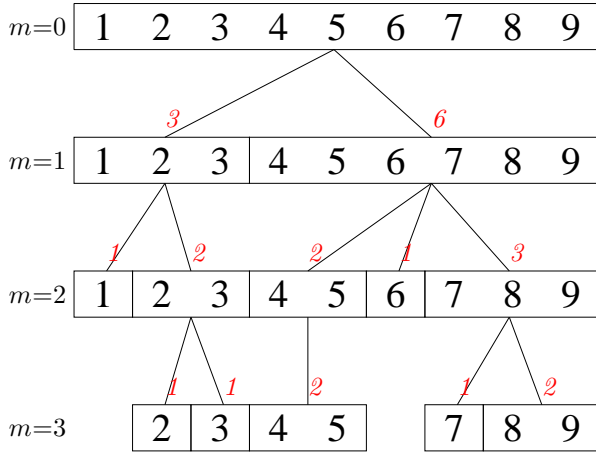


Fig. 2: An illustrative Weiqi tree with nine Weiqi games. The root vertex  $V_{0,1} = \{\mathcal{G}_{i,0} : i = 1, \dots, 9\}$  is the empty Weiqi board without any moves. After one move, at  $m = 1$ , there are two vertices,  $V_{1,1} = \{\mathcal{G}_{1,1}, \mathcal{G}_{2,1}, \mathcal{G}_{3,1}\}$  and  $V_{1,2} = \{\mathcal{G}_{4,1}, \mathcal{G}_{5,1}, \mathcal{G}_{6,1}, \mathcal{G}_{7,1}, \mathcal{G}_{8,1}, \mathcal{G}_{9,1}\}$ . After two moves at  $m = 2$ ,  $V_{1,1}$  generates two offspring vertices  $V_{2,1} = \{\mathcal{G}_{1,2}\}$  and  $V_{2,2} = \{\mathcal{G}_{2,2}, \mathcal{G}_{3,2}\}$ , while  $V_{1,2}$  generates three offspring vertices  $V_{2,3} = \{\mathcal{G}_{4,2}, \mathcal{G}_{5,2}\}$ ,  $V_{2,4} = \{\mathcal{G}_{6,2}\}$  and  $V_{2,5} = \{\mathcal{G}_{7,2}, \mathcal{G}_{8,2}, \mathcal{G}_{9,2}\}$ . Vertices  $V_{2,1}$  and  $V_{2,4}$  are leaf vertices because they contain only one game.

After the above transformations, all the openings are unique and standard, that is, any two openings  $\mathcal{G}_{i,m}$  and  $\mathcal{G}_{j,m}$  cannot overlap after further rotations and/or flips. Certainly, they might overlap without further rotations and flips. We can thus construct the Weiqi tree. Figure 2 illustrates the first four layers of the Weiqi tree with nine imaginary Weiqi games  $\{\mathcal{G}_{i,0} : i = 1, \dots, 9\}$ . In this Weiqi tree, the first layer is the root vertex  $V_{0,1}$ , which contains all nine games and corresponds to the empty Weiqi board without any moves. We denote the  $n$ -th vertices on the  $m$  layer as  $\{V_{m,i} : i = 1, 2, \dots, n\}$ . The root vertex  $V_{0,1} = \{\mathcal{G}_{i,0} : i = 1, \dots, 9\}$  is the empty Weiqi board without any moves. After one move, at  $m = 1$ , there are two vertices,  $V_{1,1} = \{\mathcal{G}_{1,1}, \mathcal{G}_{2,1}, \mathcal{G}_{3,1}\}$  and  $V_{1,2} = \{\mathcal{G}_{4,1}, \mathcal{G}_{5,1}, \mathcal{G}_{6,1}, \mathcal{G}_{7,1}, \mathcal{G}_{8,1}, \mathcal{G}_{9,1}\}$ . After two

moves at  $m = 2$ ,  $V_{1,1}$  generates two offspring vertices  $V_{2,1} = \{\mathcal{G}_{1,2}\}$  and  $V_{2,2} = \{\mathcal{G}_{2,2}, \mathcal{G}_{3,2}\}$ , while  $V_{1,2}$  generates three offspring vertices  $V_{2,3} = \{\mathcal{G}_{4,2}, \mathcal{G}_{5,2}\}$ ,  $V_{2,4} = \{\mathcal{G}_{6,2}\}$  and  $V_{2,5} = \{\mathcal{G}_{7,2}, \mathcal{G}_{8,2}, \mathcal{G}_{9,2}\}$ . Vertices  $V_{2,1}$  and  $V_{2,4}$  are leaf vertices because they contain only one game. Further,  $V_{2,2}$  splits to  $V_{3,1} = \{\mathcal{G}_{2,3}\}$  and  $V_{3,2} = \{\mathcal{G}_{3,3}\}$ ,  $V_{2,3}$  keeps unchanged to  $V_{3,3} = \{\mathcal{G}_{4,3}, \mathcal{G}_{5,3}\}$ , and  $V_{2,5}$  splits to  $V_{3,4} = \{\mathcal{G}_{7,3}\}$  and  $V_{3,5} = \{\mathcal{G}_{8,3}, \mathcal{G}_{9,3}\}$ . Each leaf vertex corresponds only to one game and cannot split. On the other hand, a vertex corresponding only to one game is always a leaf vertex.

Using the procedure described for the tree in fig. 2, we construct the Weiqi tree for all the 57 090 games. The number of vertices increases from 1 at  $m = 0$  to 11428 at  $m = 9$ , and then decreases to 518 at  $m = 30$ . By definition, the final tree has 57 090 leaf vertices. We can define the out-degree of each vertex  $V_{m,j}$ , the popularity of each opening  $\mathcal{G}_{i,m}$ , the local branching of each vertex, the uniqueness of each game, and so on. We will investigate the properties of these tree features.

**Popularity of Weiqi openings.** – The popularity  $s_{m,j}$  of an opening  $\mathcal{G}_{i,m}$  with  $m$  moves, also known as weight [19], is the strength of node  $V_{m,j}$  with  $\mathcal{G}_{i,m} \in V_{m,j}$ :

$$s_{m,j} \triangleq s(\mathcal{G}_{i,m}) \triangleq s(V_{m,j}) = |V_{m,j}|, \quad (3)$$

where  $|V|$  is the cardinality of set  $V$ . The relative popularity of an opening  $\mathcal{G}_{i,m}$  is the ratio of the popularity  $s_{m,j}$  to the total number  $N$  of all games:

$$w_{i,m} = s_{m,j}/N. \quad (4)$$

For the tree in fig. 2,  $s_{1,1} = 3$  and  $s_{1,2} = 6$  for  $m = 1$ ,  $s_{2,1} = 1$ ,  $s_{2,2} = 2$ ,  $s_{2,3} = 2$ ,  $s_{2,4} = 1$ , and  $s_{2,5} = 3$  for  $m = 2$ , and  $s_{3,1} = 1$ ,  $s_{3,2} = 1$ ,  $s_{3,3} = 2$ ,  $s_{3,4} = 1$  and  $s_{3,5} = 2$  for  $m = 3$ .

For the real tree, the first move was placed in 22 different crosses. The most popular move is  $\mathcal{G}_{m=1} = [4, 4]$  with  $s([4, 4]) = 39310$  and  $w([4, 4]) = 0.5930$ , the second most popular move is  $\mathcal{G}_{m=1} = [3, 4]$  with  $s([3, 4]) = 25350$  and  $w([3, 4]) = 0.3824$ , the third most popular move is  $\mathcal{G}_{m=1} = [3, 5]$  with  $s([3, 5]) = 743$  and  $w([3, 5]) = 0.0112$ , the fourth most popular move is  $\mathcal{G}_{m=1} = [3, 3]$  with  $s([3, 3]) = 389$  and  $w([3, 3]) = 0.0060$ , and the fifth most popular move is  $\mathcal{G}_{m=1} = [4, 5]$  with  $s([4, 5]) = 243$  and  $w([4, 5]) = 0.0037$ . Tianyuan, the central star, ranks the sixth with  $s([10, 10]) = 84$  and  $w([10, 10]) = 0.0013$ . Other openings occur much less frequently, in which four openings have 3 games ( $s(\mathcal{M}) = 3$ ) and other 7 openings appear in only one game ( $s(\mathcal{M}) = 1$ ).

For the openings  $\mathcal{G}_m$  with  $m$  moves which appear in the  $m$  layer of the Weiqi tree, we obtain the popularity  $s_m$ . Figure 3(a) illustrates the histograms of popularity,  $H(s_m)$ , for five  $m$  values.  $H(s_m)$  is the number of  $s_m$  values in each logarithmic bin. It is trivial that openings with fewer moves can have greater popularity. We find that these histogram curves have nice power-law

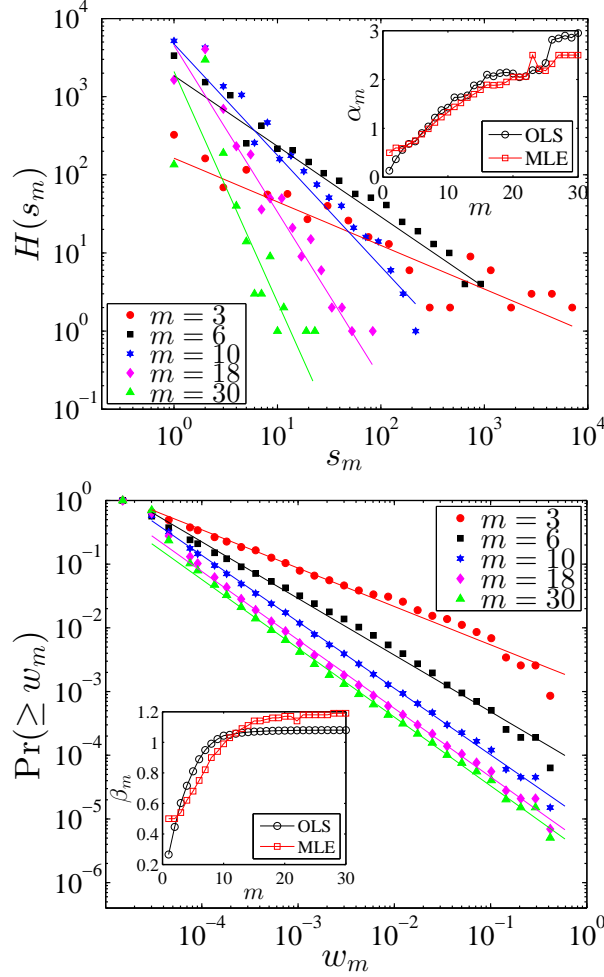


Fig. 3: (color online). (a) Histogram of popularity  $s_m$  of openings  $\mathcal{G}_m$  for  $m = 3, 6, 10, 18$ , and  $30$ . The five curves are histograms obtained with logarithmic binning. The solid lines are the best power-law fits. Inset: The slopes  $\alpha_m$  of the regression lines obtained by the ordinary linear-squares regression and the maximum likelihood estimation as a function of  $m$  for  $m = 1, \dots, 30$ . (b) Popularity distributions  $\Pr(\geq w_m)$  of pooled openings from  $\mathcal{G}_1$  to  $\mathcal{G}_m$  using logarithmic binning for  $m = 3, 6, 10, 18$ , and  $30$ . The solid lines are the best power-law fits. Inset: The slopes  $\beta_m$  of the regression lines obtained by the ordinary linear-squares regression and the maximum likelihood estimation as a function of  $m$  for  $m = 1, \dots, 30$ .

tails. The combination method of maximum likelihood estimation (MLE) and Kolgomonov-Smirnov test proposed in Ref. [20] confirms the presence of power-law tails:

$$H(s_m) \propto s_m^{-\alpha_m}, \quad (5)$$

and the tail exponents  $\alpha_m$  are obtained. We also adopt the ordinary linear-squares regression (OLS) to estimate the tail exponents. For  $m > 10$ , the first point in each curve is excluded because the first point deviates evidently from the power law. The estimated tail exponents are presented in the inset of fig. 3(a). The OLS and the MLE methods give comparable results. It is found that the exponents  $\alpha_m$

are not universal, but increases with  $m$  almost linearly. This finding for Weiqi games is consistent with that for chess openings [19].

Figure 3(b) presents the complementary cumulative distributions of the relative popularity  $\Pr(\geq w_m)$  of pooled openings from  $\mathcal{G}_1$  up to  $\mathcal{G}_m$  for  $m = 3, 6, 10, 18$ , and  $30$ . We also observe nice power-law tails in the distributions with the scaling ranges spanning approximately four orders of magnitude:

$$\Pr(\geq w_m) \propto w_m^{-\beta_m}, \quad (6)$$

which is verified by the combination method of MLE and Kolgomonov-Smirnov test in Ref. [20]. The inset shows the dependence of the tail exponents  $\beta_m$  estimated using respectively the OLS and the MLE methods on the number of moves  $m$ . Although the two  $\beta_m$  curves do not overlap, they share very a similar tendency. The tail exponent  $\beta_m$  increases with  $m$  and tends to saturation:

$$\lim_{m \rightarrow \infty} \beta_m = \beta, \quad (7)$$

where  $\beta \approx 1.08$  for the OLS and  $\beta \approx 1.19$  for the MLE. This finding implies the possibility of universality of the tail exponent, which is again consistent with the result for chess openings [19]. For chess openings, the universal exponent is reported to be “ $\alpha = 2$ ” [19], which is the power-law exponent of the density function, while the exponent  $\beta$  in Eq. (7) is the power-law exponent of the cumulative distribution. It follows that  $\beta \approx \alpha - 1$ . Our results support the conjecture that: “similar power laws could be observed in different databases and other board games, regardless of the considered game depth, constraints on player levels or the decade when the games were played” [19].

**Distribution of out-degrees.** — We can define the out-degree  $k_{m,j}$  of any vertex  $V_{m,j}$  at the  $m$ -th layer. As for the example Weiqi tree in fig. 2, we have  $k_{0,1} = 2$  for  $m = 0$ ,  $k_{1,1} = 2$  and  $k_{1,2} = 3$  for  $m = 1$ , and  $k_{2,2} = 2$ ,  $k_{2,3} = 1$  and  $k_{2,5} = 2$  for  $m = 2$ . The out-degree of leaf vertices is 0 and not included in the analysis. For the real Weiqi tree,  $k_{0,1} = 22$  as already mentioned in the previous section. When  $m = 2$ , the three largest out-degrees are 41, 30 and 28. When  $m = 29$ , the maximum out-degree reduces to 3. Overall, with the increase of  $m$ , the maximum out-degree has a decreasing trend with mild local fluctuations.

We place all the out-degrees from layer  $m = 0$  to  $m$  and estimate the underlying empirical distribution  $p(k_m)$ . Figure 4 illustrates the distributions for  $m = 3, 6, 10, 18$ , and  $30$ . When  $m$  is small, the sample is too small and the resulting distribution fluctuates a lot. The right-most point is  $\max\{k_{m=1}\} = 41$ , which appears in all distribution curves. For all the five distributions, the bulks ( $k_m > 1$ ) exhibit nice power laws:

$$p(k_m) \propto k_m^{-(\gamma_m+1)}. \quad (8)$$

The inset shows that the tail exponent  $\gamma_m$  increases when more layers are included in the opening tree.



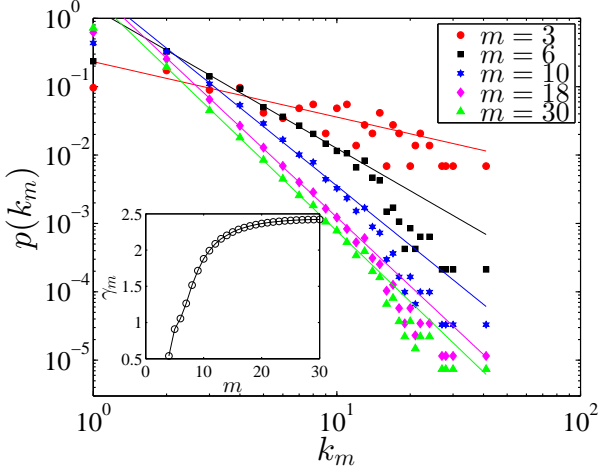


Fig. 4: (color online). Empirical out-degree distributions  $p(k_m)$  of all vertices of opening trees  $\mathcal{G}_m$  for  $m = 3, 6, 10, 18$ , and  $30$ . The solid lines are the best power-law fits. Inset: The exponents  $\gamma_m$  of the regression lines as a function of  $m$  for  $m \geq 3$ .

**Distribution of branching ratios.** — Now consider a vertex  $V_{m,j}$  that is not a leaf. Hence, we can assume that it has daughters  $V_{m+1,j_k}$  such that  $V_{m,j} = \cup_k V_{m+1,j_k}$ . We calculate the branching ratios  $r$  of the vertices as follows:

$$r_{m,j,k} = \frac{s_{m+1,j_k}}{s_{m,j}} = \frac{s(V_{m+1,j_k})}{s(V_{m,j})}. \quad (9)$$

For the artificial tree in fig. 2, there are two branching ratios  $\{1/3, 2/3\}$  for  $m = 2$ , five branching ratios  $\{1/3, 2/3, 1/3, 1/6, 1/2\}$  for  $m = 2$ , and five branching ratios  $\{1/2, 1/2, 1, 1/3, 2/3\}$  for  $m = 3$ .

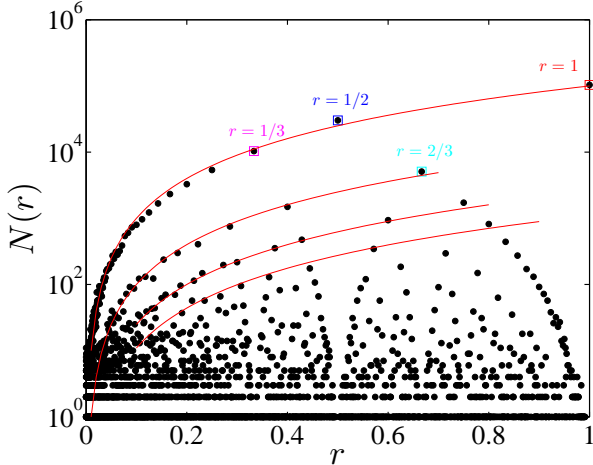


Fig. 5: (color online). Occurrence numbers of the branching ratios of non-leaf vertices of the whole Weiqi tree. The four curves are best power-law fits to four envelopes which have a form of  $N(r) \sim r^2$ .

We calculate the branching ratios of all non-leaf vertices of the Weiqi tree. The number of distinct branching ratios

is 3700. The occurrence number of each unique branching ratio is determined and presented in fig. 5. The occurrence number of  $r = 1$  is 103 484, which is the most frequent. The remaining points illustrate like a half-peach and show a hierarchical pattern. The second most frequent branching ratio is  $r = 1/2$ . Both the left and right parts with respect to  $r = 1/2$  look like half-peaches. The left part can be further divided into two half-peaches around  $r = 1/3$ , while the right part can be divided into two half-peaches around  $r = 2/3$ . This hierarchical pattern suggests an inherent self-similarity in the occurrence pattern of the branching ratios.

Intriguingly, we find that the left and right envelopes of each half-peach can be well fitted by power laws:

$$N(r) \propto \begin{cases} r^{\gamma_l}, & \text{for the left envelop} \\ -r^{\gamma_r}, & \text{for the right envelop} \end{cases} \quad (10)$$

The exponents  $\gamma_l$  and  $\gamma_r$  vary for different envelopes and are thus not universal. The four curves in fig. 5 are best power-law fits to four envelopes with  $\gamma_l = 2$ . These curves actually penetrate more points to the right of the original half-peaches.

**Distribution of uniqueness scores.** — For a given game  $\mathcal{G}_i$ , its uniqueness score  $L_i$  is the minimum number of initial moves making  $\mathcal{G}_{i,m}$  different from all other openings. Speaking differently, a game's uniqueness score is the path length from the root to the corresponding leaf. For the tree in fig. 2,  $L_1 = L_6 = 2$  and  $L_2 = L_3 = L_7 = 2$ , while  $L_3, L_4, L_8$  and  $L_9$  are longer than 3. A smaller value of uniqueness score means higher degree of uniqueness of the game.

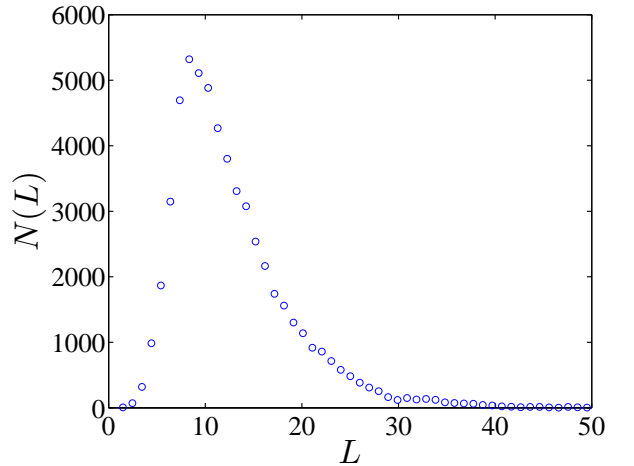


Fig. 6: (color online). Histogram of uniqueness scores of all games under investigation.

Figure 6 presents the occurrence numbers  $N(L)$  of different uniqueness scores  $L$ . The most probable uniqueness score is  $L = 9$ . The sample average is  $\langle L \rangle = 12.4$ . The proportion of games with  $5 \leq L \leq 20$  is 87.4%. These observations are due to the fact that there are wax and

wane in the popularity of Weiqi openings during different historical periods.

**Summary and discussions.** — In summary, we have proposed to construct a directed Weiqi tree based on different sequence of moves using a database of professional Weiqi games. The construction method is the same as that for the chess tree in Ref. [19].

We observed that the popularity distribution of Weiqi openings with a same number of moves is distributed according to a power law and the tail exponent increases with the number of moves. Moreover, the popularity distribution of Weiqi openings with the number of moves no more than a given number also have a power-law tail, and the tail exponent increases with the number of moves and seems to saturate to a universal value. These findings are robust using the ordinary least-squares regression and the maximum likelihood estimation in Ref. [20]. These properties are the same as for the chess tree and support the conjecture that the popularity distribution of board games follows Zipf's law with a universal exponent [19]. It has been shown in Ref. [19] that the asymptotic Zipf law arises independently from the specific form of the branching ratio distribution as well as the microscopic rules of the underlying branching process. Hence the model for the chess tree applies also to the Weiqi tree.

The distribution of out-degrees is found to have a power-law form, in which the tail exponent increases with the number of moves. We also found that the distribution of branching ratios has a very complicated pattern, in which many power laws are observed. The mechanism leading to such kind of power laws is unclear and this topic is beyond the scope of the current Letter. Finally, the distribution of uniqueness scores which define the path lengths from the root vertex to the leaf vertices has a unimodal shape and most uniqueness scores are less than 30. In other words, any two arbitrary games are very unlikely to be the same after 30 moves.

An important issue concerns the impact of sample size on the observed properties. In this vein, we need to draw a clear line between possible universality and finite-size effects if larger databases are available. We are not able to significantly increase the sample size since the number of available games is indeed limited. We thus redid the analysis using 10000 randomly chosen games from the whole sample. We find that the opening popularity distribution and the complex pattern in the branching ratio distribution are not influenced by the sample size. In contrast, both the power-law tail exponent of the out-degree distribution and the average uniqueness score decrease for the smaller sample. One needs larger databases to address the question if asymptotic features exist for these two statistical properties.

Our work provides a new angle for the study of the decision making process of Weiqi playing through a directed branching tree without loops. Some of the results reported for the Weiqi tree can be interpreted by the same mecha-

nism for the chess tree, for instance, the universal Zipf law of opening popularity. However, other empirical results such as the complicated occurrence pattern of branching ratios ask for further investigations.

\* \* \*

We are grateful to two anonymous referees for their insightful suggestions. Errors are ours. We acknowledge financial support from the Fundamental Research Funds for the Central Universities.

## REFERENCES

- [1] LICHTENSTEIN D. and SIPSER M., *J. ACM* , **27** (1980) 393.
- [2] ROBSON J. M., *The complexity of Go* in proc. of *Information Processing 83: Proceedings of IFIP Congress*, edited by MASON R. E. A., (North-Holland/IFIP, Paris) 1983 pp. 413–417.
- [3] CRĂȘMARU M. and TROMP J., *Lect. Notes Comput. Sci.* , **2063** (2002) 241.
- [4] ALLIS V., *Searching for solutions in games and artificial intelligence* Ph.d. thesis University of Limburg Maastricht, The Netherlands ISBN 90-900748-8-0 (1994).
- [5] TROMP J. and FARNEBÄCK G., *Lect. Notes Comput. Sci.* , **4630** (2007) 84.
- [6] BENSON D. B., *Inf. Sci.* , **10** (1976) 17.
- [7] CHEN K. and CHEN Z., *Inf. Sci.* , **121** (1999) 113.
- [8] VAN DER WERF E. C., WINANDS M. H., VAN DEN HERIK H. J. and UITERWIJK J. W., *Inf. Sci.* , **175** (2005) 258.
- [9] THORP E. and WALDEN W. E., *Inf. Sci.* , **4** (1972) 1.
- [10] BOUZY B. and CAZENAVE T., *Artif. Intell.* , **132** (2001) 39.
- [11] MÜLLER M., *Artif. Intell.* , **134** (2002) 145.
- [12] OSHIMA M., YAMADA K. and ENDOA S., *Procedia Computer Science* , **6** (2011) 237.
- [13] LEE B., PARK J.-Y., JUNG W. H., KIM H. S., OH J. S., CHOI C.-H., JANG J. H., KANG D.-H. and KWON J. S., *Neuroimage* , **52** (2010) 9.
- [14] LIU Z., DOU Q. and LU B., *Lect. Notes Comput. Sci.* , **5131** (2008) 125.
- [15] HARRÉ M., BOSSOMAIER T., GILLET A. and SNYDER A., *Eur. Phys. J. B* , **80** (2011) 555.
- [16] GEORGEOT B. and GIRAUD O., *EPL (Europhys. Lett.)* , **97** (2012) 68002.
- [17] HUANG S.-C., COULOM R. and LIN S.-S., *Lect. Notes Comput. Sci.* , **6515** (2011) 81.
- [18] KANDIAH V., GEORGEOT B. and GIRAUD O., *Eur. Phys. J. B* , **87** (2014) 1.
- [19] BLASIUS B. and TÖNJES R., *Phys. Rev. Lett.* , **103** (2009) 218701.
- [20] CLAUSET A., SHALIZI C. R. and NEWMAN M. E. J., *SIAM Rev.* , **51** (2009) 661.